

# Time-dependent information transmission in a model regulatory circuit

F. Mancini,<sup>1,\*</sup> C. H. Wiggins,<sup>2,†</sup> M. Marsili,<sup>3,‡</sup> and A. M. Walczak<sup>4,§</sup>

<sup>1</sup>International School for Advanced Studies (SISSA), Trieste, Italy

<sup>2</sup>Department of Applied Physics and Applied Mathematics,

Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10027

<sup>3</sup>The Abdus Salam International Centre for Theoretical Physics (ICTP), Trieste, Italy

<sup>4</sup>CNRS and Laboratoire de Physique Théorique de l'École Normale Supérieure, Paris, France.

(Dated: February 21, 2013)

Many biological regulatory systems process signals out of steady state and respond with a physiological delay. A simple model of regulation which respects these features shows how the ability of a delayed output to transmit information is limited: at short times by the timescale of the dynamic input, at long times by that of the dynamic output. We find that topologies of maximally informative networks correspond to commonly occurring biological circuits linked to stress response and that circuits functioning out of steady state may exploit absorbing states to transmit information optimally.

PACS numbers:

To respond to environmental changes, regulatory biochemical networks need to transform the molecular signals they receive as input into concentrations of response molecules. These processes are inherently stochastic, as both input and output molecules are often present in small numbers. This observation has motivated a number of recent works which pose the search for design principles as an optimization problem over the network topology and reaction rates. Such an approach has demonstrated that, within a network that functions at steady state, the system can exploit the molecular details of the network to transmit information [1–6]. Many of the current approaches have looked at instantaneous information transmission [7, 8], or the rate of information transmission [6, 9, 10]. However, regulatory response is often at a delay relative to input signaling, e.g., due to transcription and translation processes [11]. Examples include the chemotactic response of bacteria [12] or amoeba [13] to nutrients or conversely to antibiotics [14]. Optimal design therefore entails maximizing information transmission between the input at a given time and the output at a later time.

In this paper we find optimal circuit designs to transmit information in situations where the response is measured at a later time. Biochemical regulatory networks can be very complex and many of their molecular details can have an effect on their information processing functions. Here we focus on a simplified model consisting of only two binary components: the input  $z$  and the output  $x$ , which switch randomly between two discrete states,  $-1$  and  $+1$ , in continuous time (see Fig. 1 for an example time series illustrating the 8 associated transition

rates). This can model, e.g., the expression state of two genes, or a gene and a protein, each of which can be up- ( $z, x = +1$ ) or down-regulated ( $z, x = -1$ ). Already within this simple model, our approach uncovers topologies that correspond to frequently observed biological circuits [15–17].

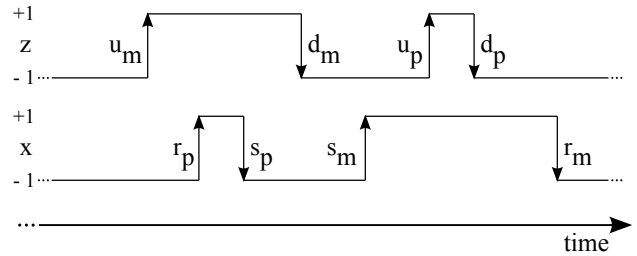


FIG. 1: Time evolution of random variable  $z(t)$ , which models a parent gene or protein transitioning from/to a down-regulated state ( $-1$ ) to/from an up-regulated state ( $+1$ ), with rates  $\{u_m, u_p\}/\{d_m, d_p\}$ , respectively. Random variable  $x(t)$  models activation ( $+1$ ) or deactivation ( $-1$ ) of a child gene (or protein): it is regulated by  $z$ , with which it aligns (‘activation’, or up-regulation) with rates  $r_m$  or  $r_p$  or anti-aligns (‘repression’, or down-regulation) with rates  $s_m$  or  $s_p$ . The subscripts  $m$  and  $p$  in the rates account for the state of the other variable, that is  $-1$  and  $+1$ , respectively.

Our system at any time is fully described by a four-state probability distribution  $p(y)$ , where  $y = (x, z) \in \{(-, -), (-, +), (+, -), (+, +)\}$  is a joint variable for the output and the input. The temporal evolution of the conditional probability  $p(y'|y)$  to find the system in state  $y'$  at time  $t$  given state  $y$  at  $t=0$  is given by a continuous-time master equation  $\partial_t p = -\mathcal{L}p$ , where  $\mathcal{L}$  is a  $4 \times 4$  transition matrix set by the rates of switching between the states (defined in Fig. 1). In  $p(y'|y)$  and in the rest of the paper, primed variables refer to the system state at time  $t$ , and unprimed variables to the initial time 0. The solution of the master equation can formally be written as  $p = e^{-t\mathcal{L}}$  and is conveniently expressed in terms of

\*fmancini@sissa.it

†chris.wiggins@columbia.edu

‡marsili@ictp.it

§awalczak@lpt.ens.fr

its expansion in left and right eigenvalues of  $\mathcal{L}$  (see Appendix A). In particular, the (normalized) right eigenvector corresponding the null eigenvalue is the stationary state  $p_\infty(y') = \lim_{t \rightarrow \infty} p(y'|y)$ .

The information between the input  $z$  at time 0 and the output  $x'$  at a later time  $t$  is defined as

$$I[x_t, z_0] = \sum_{x', z} p(x', z) \log_2 \frac{p(x', z)}{p(x')p_0(z)} \quad (1)$$

where the joint probability distribution  $p(x', z)$  can be readily derived from the conditional distribution  $p(y'|y)$  and the initial distribution  $p_0(y)$  (See Appendix A). The case in which the network must respond repeatedly to an input which is in statistical steady state differs from that in which the network must respond only once to its input, e.g. by producing an enzyme to metabolize a newly available sugar [11]. While in the former the input signal is determined by  $\mathcal{L}$  (i.e.  $p_0 = p_\infty$ ), in the latter  $p_0$  may also be optimized, as the cell may control the form of the input it presents to the regulatory network. We consider both optimization cases in the context of our model.

A trivial way to maximize information transmission is for the input to change infinitely slowly relative to a fixed delay time  $t$  (i.e.  $u_x = d_x \rightarrow 0$ ,  $s_x = 0$ ,  $r_x > 0$ ;  $x = p, m$  following the notation of Fig. 1), such that for any finite  $t$ , the output yields a noiseless readout of the input, i.e.  $p(x_t = z_0) \approx 1$ . In short, nothing happens. To constrain our optimization such that information is transduced on a timescale set by the system's own dynamics, we optimize the information

$$\mathcal{I}(\tau) = I[x_{t=\tau/\lambda}, z_0], \quad (2)$$

where the rate  $\lambda$  is given by the smallest non-zero eigenvalue of  $\mathcal{L}$ ;  $\mathcal{I}(\tau)$  implicitly depends on the rates appearing in  $\mathcal{L}$  and, if it is not set by  $p_0(y) = p_\infty(y)$ , on the initial distribution  $p_0(y)$ . We find network architectures that transmit information optimally when the response time  $t$  is comparable to the *relaxation timescale* of the system, i.e. we maximize  $\mathcal{I}$  in Eq. (2) at *fixed*  $\tau = \lambda t$  over the rates [23].

To gain intuition, we start by considering the simplest possible case (model A) where  $z$  up-regulates  $x$  perfectly, symmetrically, and without feedback. In this case,  $x$  is slaved to  $z$  and switches only if  $x \neq z$  (see Fig. 1 with  $s_m = s_p = 0$ ,  $u_m = u_p = d_p = d_m \equiv u$  and  $r_p = r_m \equiv r$ ). This leaves us with only two timescales, related to  $u$  and  $r$ .

In the steady-state case ( $p_0(y) = p_\infty(y)$ ), the mutual information can be computed explicitly and is related to the entropy of an effective two-state spin variable:

$$I[x_t, z_0] = \frac{1+\mu}{2} \log_2(1+\mu) + \frac{1-\mu}{2} \log_2(1-\mu), \quad (3)$$

where the “effective magnetization”  $\mu \equiv 2[p(x'=z, z) - p(x' \neq z, z)]$  is  $(e^{-2ut}r(r+2u) - e^{-rt}4ur)/((r-2u)(r+2u))$  (see Appendix D) [24]. For long times, we see that  $\mu \rightarrow 0$

and  $I[x_t, z_0] \rightarrow 0$ , as expected. As previously shown for a different model [18], the mutual information for a system initially in steady state has a maximum for a non-zero delay  $t^* = \log\left(\frac{2u+r}{2r}\right)/(2u-r)$ , which is determined by the interplay of the two timescales introduced above. Here, we are not interested in finding the timescales over which information transmission is maximal, but rather the rates that maximize  $\mathcal{I}(\tau)$  at *fixed* rescaled time [25]  $\tau = \lambda t$ , where  $\lambda = \min\{2u, r\}$  for model A (see Fig. 5).

The optimal information  $\mathcal{I}^*(\tau)$  and parameters ( $u^*(\tau), r^*(\tau)$ ) for the simplest model where  $z$  activates  $x$  are plotted in Fig. 2 A [26]. We see a clear crossover in terms of the parameter  $u^*$  that regulates the state of  $z$  (see dashed vertical line in Fig. 2 A): it initially increases in time and then plateaus at  $u^*=0.5$ . This crossover results from the fact that for  $r > 2u$  the relaxation time is dominated by the rate at which the input changes ( $\lambda=2u$ ), whereas for  $r \leq 2u$  the rate at which the output changes fixes relaxation times ( $\lambda=r$ ). Information transmission is dominated, over short timescales, by the faster rate  $r$ . Over long timescales, optimality is achieved by matching the characteristic times of the two processes ( $r=2u$ ). The degeneracy of the two smallest non-zero eigenvalues is a non-trivial generic feature of optimal networks that we also find in more complex models (see below). The dynamics of model A can be summarized by the network topology and regulatory circuits shown in Fig. 3 A.

We can generalize model A by allowing  $z$  to down-regulate  $x$  as well — that is, to allow  $s_m = s_p \neq 0$  (model B). As in model A, we forbid feedback from  $x$  to  $z$ , hence the transition rates for  $z$  do not depend on the state of  $x$  (i.e.  $u_m = u_p = u$  and  $d_p = d_m = d$ ). Optimization yields only solutions coinciding with that of model A, or with its symmetric counterpart (wherein  $z$  perfectly down-regulates  $x$  instead of perfectly up-regulating:  $r_p = r_m = 0$  and  $s_p = s_m = s$ ). Intuitively, in order for information to be transduced between  $x$  and  $z$ , they either align or anti-align, resulting in the common simple activator or repressor element [19]. Note that the same topological structure is found at all timescales  $\tau$ . This is to be contrasted with previous studies [2, 20] which, taking into account the molecular cost paid by producing higher copy number (e.g., creating more proteins), have found small discrepancies between the information transmitted in the two cases of up-regulation and down-regulation. Since our model does not explicitly account for protein copies, we do not observe such a difference.

Recent studies [6, 10, 21] have pointed to the important role of feedback in transmitting information, a form of which we can consider using the full set of 8 rates in Fig. 1. Now the hierarchical relation between  $z$  and  $x$  is broken: both can regulate each other's expression, either by down- or up-regulation. Examples of maximally informative topologies for all possible time delays  $\tau$  are illustrated in Fig. 3 C and reveal a “push-pull” network — one gene (or protein) up-regulates the other, which in turn down-regulates the first gene/protein. Again, due to

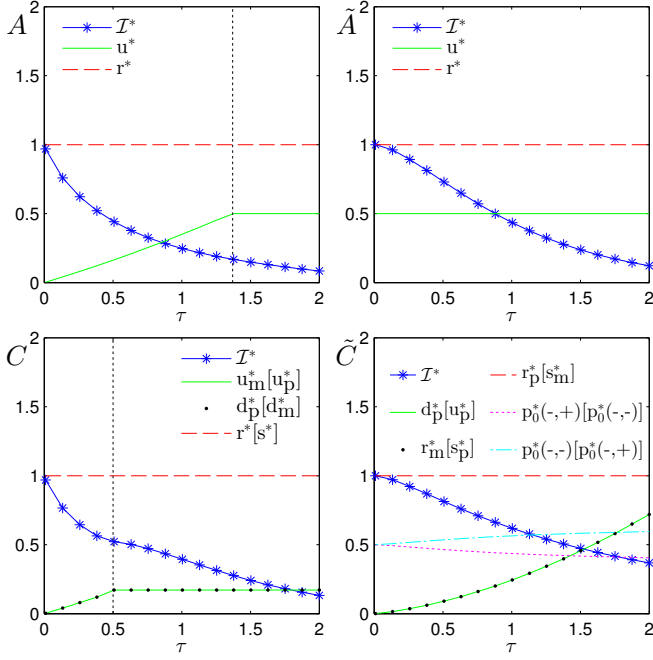


FIG. 2: Optimal parameters and transmitted information  $T^*$  for the activator and feedback models in steady state ( $A$  and  $C$ ) and relaxing the steady state assumption ( $\tilde{A}$  and  $\tilde{C}$ ), as functions of the rescaled time  $\tau$ . In panels  $C$  and  $\tilde{C}$ , parameters in square brackets refer to alternative optimal topologies (see right diagrams in Fig. 3  $C$  and  $\tilde{C}$ ). Subscripts  $m, p$  are omitted when  $x_m = x_p$  (with  $x = u, d, r, s$ ). Results for models  $B$  and  $\tilde{B}$  are shown in Appendix G.

the symmetry of the problem, we can flip either  $x$  or  $z$  and we find 4 equally informative solutions (two of which are shown in Fig. 3  $C$ ), associated with different sets of the 8 rates being driven to zero by the optimization procedure. The information transmission is again controlled by the relaxation time, which switches from being dominated by the input timescale at small times, to being dominated by the output timescale at longer times (see dashed vertical line in Fig. 2  $C$ ). We find that the optimal value of  $u$  now plateaus at  $3 - 2\sqrt{2}$  for long times: this value is set by competition with the  $r$  rate, by matching the two smallest non-zero eigenvalues in order to avoid oscillatory solutions (For  $u > u^* = 3 - 2\sqrt{2}$ , the eigenvalues become imaginary describing oscillations; see Appendix E for a more detailed derivation). Push-pull networks can oscillate [22] [27], thwarting optimal information transmission by decorrelating the system; the optimal parameters instead avoid oscillations, such that one of the regulatory timescales is much less than the other. This increases the relaxation time of the system and the circuit in model  $C$  can transmit more information at a fixed relaxation time than the optimal circuit without feedback (Fig. 2). Feedback leads to a rotational directionality among the transitions (cf. Fig. 3), in which the state never ‘flips back’, enhancing the transduced information.

Such push-pull circuits are very common in biol-

ogy, from microbes [15] to humans ([16] and references therein) as a source of oscillations [22] and pulse responses [16] (see discussion below).

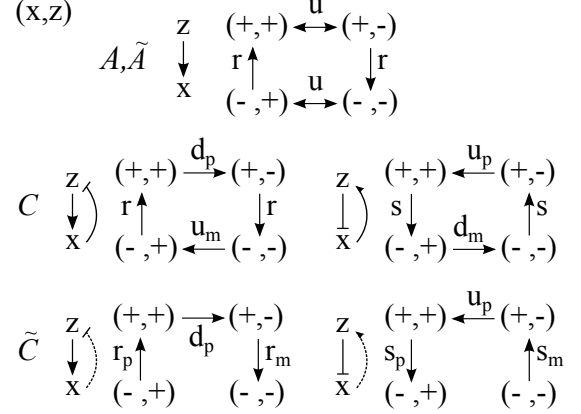


FIG. 3: Optimal topologies and dynamics for the activator and feedback models in steady state ( $A$  and  $C$ ) and relaxing the steady state assumption ( $\tilde{A}$  and  $\tilde{C}$ ). In the two feedback models ( $C$  and  $\tilde{C}$ ) the optimal topology is a ‘push-pull’: one gene (or protein) activates the other, which in turn represses the first gene/protein (as we can see, the roles of  $z$  and  $x$  are interchangeable). The dashed line in  $\tilde{C}$  means that the feedback exists only until the absorbing state is reached. The topologies of these optimal networks were found by inspecting the optimal rates manually. Results for models  $B$  and  $\tilde{B}$  are shown in Appendix G.

Having discussed optimal delayed information transmission of repeated readouts in the stationary state, we now consider regulatory networks that optimize a one time response to an external signal (e.g. by producing an enzyme to metabolize a nutrient that appeared in the environment). Specifically, we consider the same three models studied above ( $A$ ,  $B$ ,  $C$ ), but now optimize simultaneously not only on the rates but also on the initial probability distribution  $p_0(y)$  and refer to the associated models as  $\tilde{A}$ ,  $\tilde{B}$  and  $\tilde{C}$ , respectively. We start with model  $\tilde{A}$ , which enjoys an up-down symmetry and suggests parameterizing the initial distribution  $p_0(y) = p_0(x, z)$  via  $p_0(+, +) = p_0(-, -) = (1 + \mu_0)/4$  and  $p_0(+, -) = p_0(-, +) = (1 - \mu_0)/4$  [28]. The mutual information is still given by Eq. 3, with  $\mu = \mu_0 e^{-rt} + \frac{r}{r-2u}(e^{-2ut} - e^{-rt})$ . For  $t=0$ , the highest information is attained for  $\mu = \mu_0 = \pm 1$ , when  $z$  and  $x$  are perfectly aligned/anti-aligned. Moreover,  $\mu$  and  $I[x_t, z_0]$  decrease exponentially with time  $t$ . Therefore, unlike in model  $A$ , the information transmission does not improve by making a delayed readout.

After performing the optimization, we find that for each timescale  $\tau$  the optimal initial state concentrates on the states  $(+, +)$  and  $(-, -)$ , so that  $I[x_0, z_0] = 1$ . The absence of a maximum at  $t^* > 0$  in  $I[x_t, z_0]$  for optimal initial states (model  $\tilde{A}$ ) suggests that, at odds with the stationary case (model  $A$ ), the mechanism for information transmission is only governed by the loss of informa-

tion about the initial state as it relaxes to steady state.

When information transmission is maximized over the initial distribution in the more general models  $\tilde{B}$  and  $\tilde{C}$ , we find a qualitative difference in design as compared to the steady state case (models  $B$  and  $C$ ): while the optimal topology remains the same, now either one of the aligned or non-aligned states becomes an absorbing state, e.g.,  $p_\infty(y') = \delta_{y',(+,+)}$  or  $p_\infty(y') = \delta_{y',(-,+)}$  in the examples in Fig. 3 $\tilde{C}$  [29]. As above, symmetry provides a number of optimal networks related by permutations (see Fig. 7 $\tilde{B}$  and 8); the rates of the optimal networks are shown for a case of  $z$  up-regulating [down-regulating]  $x$  in Fig. 6  $\tilde{B}$  and Fig. 2 $\tilde{C}$ .

The occurrence of an absorbing state, with a nearly-equal initial distribution over the initial and final states, limits the accessible states and leads to the optimal topology for a one time response. In the absence of feedback (model  $\tilde{B}$ , e.g. receptor activation in a complex pathway), when the system, initially in the inactive state  $(x, z) = (-, -)$  is presented with a signal  $(x, z) \rightarrow (-, +)$ , it switches on a response  $(x, z) \rightarrow (+, +)$  (see Fig. 7 $\tilde{B}$ ). However, in the presence of feedback (model  $\tilde{C}$ , e.g. a nutrient activating the production of an enzyme for its uptake, amino acid biosynthesis) the optimal dynamics includes “feedback inhibition”, in which the output switches off the input [11][30]. As in model  $C$ , feedback imposes an order to the visited states, with a smaller rate for  $z$  transitions than for  $x$  transitions, and the initial distribution admits only two initial states  $((-, +), (-, -))$  in Fig. 3 $\tilde{C}$  left), so that  $I[x_0, z_0] = 1$ , as in model  $\tilde{A}$ . Inspecting Fig. 3 $\tilde{C}$  left, and recalling that the rate for  $z$  transitions is small, we see that  $x_t = +$  could only imply  $z_0 = +$ ; similarly,  $x_t = -$  means the system started in the absorbing state  $(-, -)$ . In the biological example above, this corresponds to the enzyme switching off  $(-, -)$  until a new source of sugar appears.

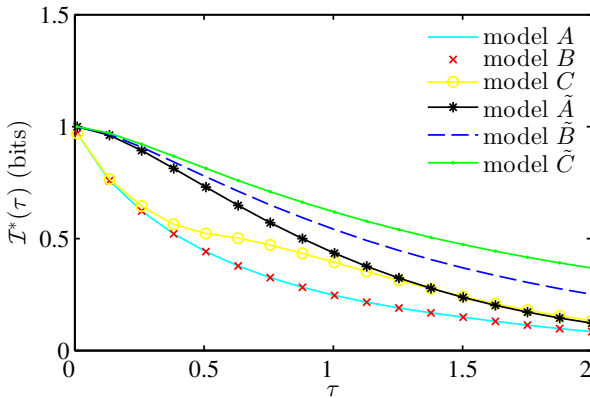


FIG. 4: Summary of optimization results: for each of the six models, the hull curve of the maximized information  $I(\tau)$  is plotted versus  $\tau$ . The information decay is slower when feedback is present (model  $C$ ) and when the system is initially in an optimal state (model  $\tilde{A}$ ,  $\tilde{B}$ ,  $\tilde{C}$ ).

In Fig. 4 we plot a comparison of all the cases possi-

ble within the two-state model. As the model generality increases (from  $A$  to  $\tilde{C}$ ), so does the number of parameters; accordingly, the information capacity of the system also increases. We see that the introduction of feedback in model  $C$  does not play a role in increasing information transmission for small  $\tau$  (input-switching dominated regime). However, the information gain coming from feedback is substantial for long time delays  $\tau$  between the input and output readout (output-switching dominated regime). Information transmission between the input and the output can be improved beyond that achieved in the steady state ( $A$ ,  $B$  and  $C$ ) if the system is pushed out of equilibrium in specific ways ( $\tilde{A}$ ,  $\tilde{B}$  and  $\tilde{C}$ ) to respond to one time signals. The information gain is particularly significant at fast time-scales.

Push-pull circuits exists in many cells, ranging from bacterial (heat-shock response [15]) to mammalian (I- $\kappa$ B-NF $\kappa$ B circuit [17] in animal stress response, p53-Mdm2 network involved in DNA damage response [16]), and often combine a slow (transcriptional regulation) with a fast (protein-protein interaction) component, similarly to the design of our optimal architectures, resulting in pulse like responses to stress signals. Both in the steady state and the non-steady state response we find the same network topology of a push-pull network. The difference between the designs is the absorbing state, which results in a single pulse response. The following pulse must be triggered by a new signal, hinting a digital type of response [16].

This study gives a framework for studying information transmission between two time points in biochemical regulatory systems. Such an approach can be extended to more realistic models that explicitly account for protein concentrations, where costs of protein production and degradation can be studied in detail.

**Acknowledgments.** We thank William Bialek for helpful discussions. AMW is funded by a Marie Curie Career Integration Grant.

## Appendix A: Calculating mutual information

In the main text we calculate the mutual information between the input  $z$  at time 0 and the output  $x'$  at a later time  $t$  using the temporal evolution of the joint probability distribution  $p(x', z)$  obtained from a master equation. In this Appendix we give a detailed derivation of the steps of this calculation.

We define the state  $y$  of the system as

$$y = (x, z) \in \{(-, -), (-, +), (+, -), (+, +)\}, \quad (\text{A1})$$

and the dynamics in terms of transition rate matrix

$$\mathcal{L} = \begin{pmatrix} u_m + s_m & -d_m & -r_m & 0 \\ -u_m & d_m + r_p & 0 & -s_p \\ -s_m & 0 & u_p + r_m & -d_p \\ 0 & -r_p & -u_p & d_p + s_p \end{pmatrix}. \quad (\text{A2})$$

The corresponding master equation is

$$\frac{dp}{dt} = -\mathcal{L}p \quad (\text{A3})$$

for a vector  $p=p(y)=(p(-,-), p(-,+), p(+,-), p(+,+))$  (we omit the implied dependence on time). Primed variables (e.g.,  $y'$ ) refer to the state of the system at time  $t \neq 0$ ; unprimed ones refer to the state at  $t=0$ . The transition probability matrix  $p(y'|y)$  is a solution of the master equation with initial condition

$$\lim_{t \rightarrow 0} p(y'|y) = \delta_{y',y}. \quad (\text{A4})$$

It can be written as the  $(y', y)$  element of the matrix  $e^{-t\mathcal{L}}$ , i.e.

$$p(y'|y) = [e^{-t\mathcal{L}}]_{y',y} = \sum_{\alpha=1}^4 e^{-\lambda_{\alpha}t} v_{\alpha}(y') u_{\alpha}(y), \quad (\text{A5})$$

where  $\lambda_{\alpha}$  (with  $\alpha = 1, \dots, 4$ ) are the four (assumed to

be distinct for this derivation) eigenvalues of  $\mathcal{L}$ , and  $v_{\alpha}$  and  $u_{\alpha}^T$  are their corresponding orthonormal right and left eigenvectors, with components  $v_{\alpha}(y')$  and  $u_{\alpha}(y)$ :

$$\mathcal{L}v_{\alpha} = \lambda_{\alpha}v_{\alpha}, \quad (\text{A6})$$

$$u_{\alpha}^T \mathcal{L} = u_{\alpha}^T \lambda_{\alpha}, \quad (\text{A7})$$

$$u_{\alpha}^T v_{\beta} = \delta_{\alpha,\beta}. \quad (\text{A8})$$

In particular, if we choose a normalization such that  $u_1 = (1, 1, 1, 1)$ , the eigenvector  $v_1$ , corresponding to the eigenvalue  $\lambda_1 = 0$ , is the stationary state  $p_{\infty}(y)$ .

We are computing the mutual information between  $z$  at time 0 and  $x'$  at time  $t$ , which is a function of  $t$  and is given by

$$I[x_t, z_0] = \sum_{x',z} p(x', z) \log_2 \frac{p(x', z)}{p(x')p(z)}. \quad (\text{A9})$$

---

The joint probability  $p(x', z)$  is calculated from  $p(y'|y)$  as

$$\begin{aligned} p(x', z) &= \sum_{y,y'} p(x', z|y', y) p(y', y) && \text{using the definition of conditional probabilities} \\ &= \sum_{y,y'} p(x'|y') p(z|y) p(y', y) && \text{exploiting the conditional independence of } x', z \\ &= \sum_{y,y'} p(x'|y') p(z|y) p(y'|y) p_0(y) && \text{using the definition of conditional probabilities.} \end{aligned} \quad (\text{A10})$$

Note that the elements of  $p(x'|y')$  and  $p(z|y)$  are either 0 or 1 according to whether, for example,  $y$  is consistent or inconsistent with  $z$ :

$$\begin{aligned} p(z = +|y = (+, +)) &= 1 \\ p(z = +|y = (+, -)) &= 0 \\ &\vdots \end{aligned} \quad (\text{A11})$$

et cetera. Finally, the marginal probabilities  $p(z)$  and  $p(x')$  are given by

$$p(z) \equiv \sum_{x'} p(x', z), \quad p(x') \equiv \sum_z p(x', z).$$

The numerical computation of the mutual information can now be implemented and the optimal rates for systems of various complexity can be found numerically. In the paragraphs below we present useful computational details for implementing this calculation in MATLAB. We have implemented the optimization procedure in MATLAB and we made the source code available via the following public access repository: <http://infodyn.sourceforge.net>.

For certain models we can also make analytical progress by exploiting spectral representations of the

joint distribution, as shown in the paragraphs below; we present these results in Appendices D, E and F.

**Numerical computation of the joint distribution.** For numerical computation in MATLAB, it is useful to rewrite Eqn. A10 in terms of matrix operations. To that end, we define (note that  $X$  and  $Z$  are 0 – 1 matrices – whose elements are 0 or 1, as per Eqn. A11)

$$X_{x',y'} \equiv p(x'|y') \quad (\text{A12})$$

$$G_{y',y_0} \equiv p(y'|y_0) = [e^{-t\mathcal{L}}]_{y',y_0} \quad (\text{A13})$$

$$P_{y_0,y} \equiv p_0(y_0) \delta_{y_0,y} \quad (\text{A14})$$

$$Z_{y,z} \equiv p(z|y). \quad (\text{A15})$$

This allows us to write Eqn. A10 compactly as

$$p(x', z) = \sum_{y'} X_{x',y'} \sum_{y_0} G_{y',y_0} \sum_y P_{y_0,y} Z_{y,z} \quad (\text{A16})$$

$$= [XGPZ]_{x',z} \quad (\text{A17})$$

that is how the equation is implemented in MATLAB.

**Analytical calculation of the joint distribution.** For analytic calculations, it is useful to expand  $p(y'|y)$  in Eqn. A10 in terms of its spectral representation (Eqn. A5):

$$\begin{aligned}
p(x', z) &= \sum_{\alpha} \sum_{y', y} p(x'|y') p(z|y) e^{-\lambda_{\alpha} t} v_{\alpha}(y') u_{\alpha}(y) p_0(y) \\
&= \sum_{\alpha} e^{-\lambda_{\alpha} t} \left( \sum_{y'} p(x'|y') v_{\alpha}(y') \right) \left( \sum_y p(z|y) u_{\alpha}(y) p_0(y) \right) \\
&= \left( \sum_{y'} p(x'|y') p_{\infty}(y') \right) \left( \sum_y p(z|y) p_0(y) \right) + \sum_{\alpha > 1} e^{-\lambda_{\alpha} t} \left( \sum_{y'} p(x'|y') v_{\alpha}(y') \right) \left( \sum_y p(z|y) u_{\alpha}(y) p_0(y) \right) \\
&\equiv p_{\infty}(x') p_0(z) + \sum_{\alpha > 0} e^{-\lambda_{\alpha} t} \tilde{v}_{\alpha}^{x'} \tilde{u}_{\alpha}^z
\end{aligned} \tag{A18}$$

where

$$p_{\infty}(x') \equiv \sum_{y'} p(x'|y') p_{\infty}(y') \tag{A19}$$

$$p_0(z) \equiv \sum_y p(z|y) p_0(y) \tag{A20}$$

$$\tilde{v}_{\alpha}^{x'} \equiv \sum_{y'} v(y')_{\alpha} p(x'|y') \tag{A21}$$

$$\tilde{u}_{\alpha}^z \equiv \sum_y u_{\alpha}(y) p_0(y) p(z|y). \tag{A22}$$

Writing  $p(x', z)$  in this form makes it clear that, if the eigenvalues are distinct and thus  $p(y'|y)$  is diagonalizable, then  $p(x', z)$  factorizes as  $t \rightarrow \infty$  and thus  $I[x_t, z_0] \rightarrow 0$ . Also clear is that  $p(x', z) - p_{\infty}(x') p_0(z)$  is expressible as a sum of time-decaying exponentials. Since  $p(x'|y')$  and  $p(z|y)$  are 0 – 1 matrices, in many cases  $\tilde{v}_{\alpha}$  and  $\tilde{u}_{\alpha}$  can be calculated explicitly, as shown below.

## Appendix B: Optimization procedure

We optimize over the parameters of each model in order to maximize  $\mathcal{I}(\tau) = I[x_{t=\tau/\lambda}, z_0]$ , where  $\tau$  is a dimensionless quantity that results from the rescaling procedure:

$$t \rightarrow t \cdot \lambda \equiv \tau,$$

where  $\lambda$  is the system's relaxation rate (the smallest nonzero eigenvalue of the rate matrix  $\mathcal{L}$ ). The steps of the “rescale and optimize” procedure are:

- while  $\tau_{min} < \tau < \tau_{max}$ :
  1. optimize  $\mathcal{I}(\tau; \theta)$  over parameters  $\theta$  or parameters  $\theta$  and initial distribution  $p(y)$
  2. save  $\mathcal{I}^*, \theta^*, p_0^*$
  3. increment  $\tau$
- end loop over  $\tau$

where

calculate  $\mathcal{I}(\tau; \theta)$ :

1. calculate  $\mathcal{L}(\theta)$
  2. calculate  $\lambda(\mathcal{L})$
  3. calculate  $p(x, z) = X \exp(-\tau \mathcal{L} / \lambda) P Z$ , as in Eqn. A17
  4. calculate  $I[p(x, z)]$
- return  $I[p(x, z)]$  to optimization algorithm

The results are obtained as hull plots, as presented in Fig. 5 for model A.

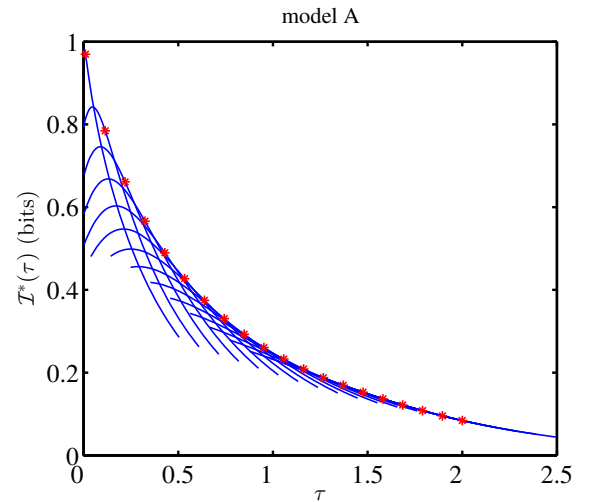


FIG. 5: An explicit construction of the optimal information curves presented in the main text, here shown for model A. The maximum value  $\mathcal{I}^*(\tau)$  for each  $\tau$  (red \*) is obtained by optimizing the rates  $u$  and  $r$  at each  $\tau$ . The optimal rates can be different for each  $\tau$ . The blue continuous curves show the whole range of  $\mathcal{I}(\tau)$ : each of them intersects a  $(\tau_{\mathcal{I}^*}, \mathcal{I}^*(\tau))$  and is computed by using the corresponding optimal set of rates.

### Appendix C: Derivation of mutual information for 2-bit, symmetric systems

In this Appendix we derive useful relations for the mutual information in the case of 1-bit symmetric systems. Consider any arbitrary distribution  $p(a, b)$  where  $\{a, b\} \in \{-1, +1\}$  and the distribution enjoys symmetry under flipping  $-1 \leftrightarrow +1$ . In this case  $p(+, +) = p(-, -)$  and  $p(+, -) = p(-, +)$ . For any such distribution the mutual information greatly simplifies, as exploited in the analytic results associated with model A.

Let us define  $-1 < \eta < 1$  such that

$$p(+, +) = p(-, -) = (1 + \eta)/4, \quad (C1)$$

$$p(-, +) = p(+, -) = (1 - \eta)/4. \quad (C2)$$

From these, we see that

$$p(a) = p(b) = 1/2 \quad (C3)$$

and

$$\begin{aligned} I[a, b] &= \sum_{a, b} p(a, b) \log_2 p(a, b) / (p(a)p(b)) \\ &= \sum_{a, b} p(a, b) \log_2 4p(a, b) \\ &= 2p(-, -) \log_2 4p(-, -) + 2p(-, +) \log_2 4p(-, +) \\ &= \frac{1}{2}(1 + \eta) \log_2(1 + \eta) + \frac{1}{2}(1 - \eta) \log_2(1 - \eta) \end{aligned} \quad (C4)$$

as in the main text, where we replace  $\eta$  by the appropriate expression for  $\mu$  in each model.

Note also that

$$\begin{aligned} p(a = b, b) - p(a \neq b, b) &= \\ p(+, +) - p(-, +) &= \\ 1/4(1 + \eta - 1 + \eta) &= \eta/2. \end{aligned} \quad (C5)$$

### Appendix D: Explicit calculation for model A

Model A describes a system in which  $z$  up-regulates  $x$  and  $x$  is slaved to  $z$ . In this case we can diagonalize  $\mathcal{L}$  analytically and calculate the mutual information. We set

- $u_m = u_p = d_m = d_p = u$ ,
- $r_m = r_p = r = 1$ ,
- $s_m = s_p = 0$ ,

and the initial probability  $p_0(y)$  is set equal to the steady state  $p_\infty(y)$ .

In this case the transition rate matrix  $\mathcal{L}$  is given by

$$\mathcal{L} = \begin{pmatrix} u & -u & -r & 0 \\ -u & u+r & 0 & 0 \\ 0 & 0 & u+r & -u \\ 0 & -r & -u & u \end{pmatrix}$$

and its spectrum is

$$\{\lambda_1 = 0, \lambda_2 = r, \lambda_3 = 2u, \lambda_4 = r + 2u\}. \quad (D1)$$

The relaxation rate of the system is given by the first non-zero eigenvalue, which switches from  $\lambda_{\text{relaxation}} = \lambda_3 = 2u$  for small  $\tau$  to  $\lambda_{\text{relaxation}} = \lambda_2 = r$  for large  $\tau$ . This change in the rates that govern the relaxation times marks the crossover shown in Fig. 2 in the main text.

The right eigenvectors are

$$\begin{aligned} v_1 &= \frac{1}{2(r+2u)} \begin{pmatrix} r+u \\ +u \\ +u \\ r+u \end{pmatrix}, v_2 = \frac{1}{2(r-2u)} \begin{pmatrix} +u \\ +u \\ -u \\ -u \end{pmatrix}, \\ v_3 &= \frac{1}{2(r-2u)} \begin{pmatrix} -r+u \\ -u \\ +u \\ +r-u \end{pmatrix}, v_4 = \frac{1}{2(r+2u)} \begin{pmatrix} +u \\ -u \\ -u \\ +u \end{pmatrix}. \end{aligned} \quad (D2)$$

The left eigenvectors are

$$\begin{cases} u_1^T = (1, 1, 1, 1), \\ u_2^T = (-1, \frac{-u+r}{u}, \frac{+u-r}{u}, 1), \\ u_3^T = (-1, 1, -1, 1), \\ u_4^T = (1, \frac{-u-r}{u}, \frac{-u-r}{u}, 1). \end{cases} \quad (D3)$$

Using the expressions introduced in Appendix C, we find that  $p(x', z)$  is given by the following  $2 \times 2$  matrix:

$$p(x', z) = \begin{pmatrix} \frac{(1+e^{-2tu})r^2+2(-2e^{-rt}+e^{-2ut})ru-4u^2}{4(r-2u)(r+2u)} & \frac{(1-e^{-2ut})r^2+2(+2e^{-rt}-e^{-2ut})ru-4u^2}{4(r-2u)(r+2u)} \\ \frac{(1-e^{-2ut})r^2+2(+2e^{-rt}-e^{-2ut})ru-4u^2}{4(r-2u)(r+2u)} & \frac{(1+e^{-2ut})r^2+2(-2e^{-rt}+e^{-2ut})ru-4u^2}{4(r-2u)(r+2u)} \end{pmatrix}. \quad (D4)$$

We can then compute  $I[x_t, z_0]$ , which, after some algebraic manipulation, reads:

$$I[x_t, z_0] = \frac{1}{2} \left( 1 + \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)} \right) \log_2 \left[ 1 + \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)} \right] + \frac{1}{2} \left( 1 - \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)} \right) \log_2 \left[ 1 - \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)} \right]. \quad (D5)$$

If we introduce the quantity

$$\mu = \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)}, \quad (D6)$$

we recover Eqn. 3 in the main text. We note that for large  $\tau$  the optimal rates are  $r^*=2u^*$ , so that  $\lambda_{\text{relaxation}}=\lambda_2=\lambda_3$ . For this special case, the expressions above may be evaluated by Taylor expanding about  $r=2u$  to find

$$\mu = e^{-rt}/2(1-2rt). \quad (D7)$$

We see that for  $t=0$ ,  $I[x_t, z_0]=1$  bit and, for long times,  $\mu \rightarrow 0$  and  $I[x_t, z_0] \rightarrow 0$ .

Moreover, on all time-scales we find that in the stationary state  $p(x'=z)/p(x'=-z)=(r/u+1) > 1$ .

From Eqn. D5, taking  $\frac{\delta I}{\delta t}|_{t^*}=0$  and using the appropriate expression for  $\mu$ , we are able to find the optimal time lag for model *A* as

$$t^* = \log \left( \frac{2u+r}{2r} \right) / (2u-r) \text{ for } r \neq 2u, \quad (D8)$$

or  $t^* = \frac{1}{2r}$ , for  $r = 2u$ , as presented in the main text.

## Appendix E: Extension to model *C*

In model *C* we allow for all the eight rates to be nonzero and different from each other. Via the symmetries of the

system (e.g., relabeling the nodes and their rates), there are many parameter settings which perform identically. Qualitatively, these topologies may all be described as either

1.  $z$  activates  $x$ , which in turn represses  $z$ , or
2.  $z$  represses  $x$ , which in turn activates  $z$ .

As an example, we consider a topology of the first type, where the rates are the following:

- $u_m = d_p$ ,
- $u_p = d_m$ ,
- $r_m = r_p$ ,
- $s_p = s_m = 0$ .

The spectrum of  $\mathcal{L}$  is

$$\begin{cases} \lambda_1 = 0, \\ \lambda_2 = r_m + u_m, \\ \lambda_3 = \frac{1}{2} \left( r_m + u_m - \sqrt{r_m^2 - 6r_mu_m + u_m^2} \right), \\ \lambda_4 = \frac{1}{2} \left( r_m + u_m + \sqrt{r_m^2 - 6r_mu_m + u_m^2} \right). \end{cases} \quad (E1)$$

For small  $\tau$  the relaxation time is given by  $\lambda_3$ . For  $u_m = (3-2\sqrt{2})r_m$ ,  $\lambda_3 = \lambda_4$ . Since we constrain the timescales such that  $\max\{u_m, d_p, u_p, d_m, r_m, r_p\} = 1$ , for large  $\tau$  the optimal value of  $r_m$  and  $u_m$  are  $r_m^* = 1$  and  $u_m^* = 3-2\sqrt{2} < 1$ , as can be seen from Fig. 2*C*. The right eigenvectors are

$$v_1 = \frac{1}{2(r_m+u_m)} \begin{pmatrix} r_m \\ u_m \\ u_m \\ r_m \end{pmatrix}, v_2 = \frac{1}{2(r_m+u_m)} \begin{pmatrix} +u_m \\ -u_m \\ -u_m \\ +u_m \end{pmatrix}, \quad (E2)$$

$$v_3 = \begin{pmatrix} \frac{-r_m+u_m-\sqrt{r_m^2-6r_mu_m+u_m^2}}{4\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{-u_m}{2\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{+u_m}{2\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{+r_m-u_m+\sqrt{r_m^2-6r_mu_m+u_m^2}}{4\sqrt{r_m^2-6r_mu_m+u_m^2}} \end{pmatrix}, v_4 = \begin{pmatrix} \frac{+r_m-u_m-\sqrt{r_m^2-6r_mu_m+u_m^2}}{4\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{+u_m}{2\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{-u_m}{2\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{-r_m+u_m+\sqrt{r_m^2-6r_mu_m+u_m^2}}{4\sqrt{r_m^2-6r_mu_m+u_m^2}} \end{pmatrix}. \quad (E3)$$



The left eigenvectors are

$$\begin{cases} u_1^T = (1, 1, 1, 1), \\ u_2^T = (1, -\frac{r_m}{u_m}, -\frac{r_m}{u_m}, 1), \\ u_3^T = (-1, \frac{+r_m - u_m - \sqrt{r_m^2 - 6r_mu_m + u_m^2}}{2r_m}, \frac{-r_m + u_m + \sqrt{r_m^2 - 6r_mu_m + u_m^2}}{2r_m}, 1), \\ u_4^T = (-1, \frac{+r_m - u_m + \sqrt{r_m^2 - 6r_mu_m + u_m^2}}{2r_m}, \frac{-r_m + u_m - \sqrt{r_m^2 - 6r_mu_m + u_m^2}}{2r_m}, 1). \end{cases} \quad (\text{E4})$$

The mutual information is given by Eqn. 3 in the main text, with  $\mu$  given by

$$\mu = \cosh(t\sqrt{u_m^2 + r_m^2 - 6r_mu_m}) \frac{r_m - u_m}{r_m + u_m} + \sinh(t\sqrt{u_m^2 + r_m^2 - 6r_mu_m}) \frac{u_m^2 - r_m^2 - 4r_mu_m}{(r_m + u_m)(u_m^2 + r_m^2 - 6r_mu_m)}. \quad (\text{E5})$$

For  $t=0$ ,  $\mu_0 = \frac{r_m - u_m}{r_m + u_m}$  and  $I[x_t, z_0] = 1$  bit, as can be seen from Fig. 2  $\tilde{C}$  in the main text. We can define

$$x = t\sqrt{u_m^2 + r_m^2 - 6r_mu_m} \quad (\text{E6})$$

and, noting that

$$\lim_{x^* \rightarrow 0} \frac{\sinh x^*}{x^*} = 1, \quad (\text{E7})$$

we obtain for large  $\tau$  the optimal

$$\mu^* = \frac{r_m - u_m}{r_m + u_m} + \frac{u_m^2 - r_m^2 - 4r_mu_m}{r_m + u_m} t. \quad (\text{E8})$$

## Appendix F: Model $\tilde{A}$

Model  $\tilde{A}$  is the same as model  $A$ , except for the fact that the initial probability  $p_0(y)$  is not given by the steady state but is optimized. In order to optimize  $I[x_t, z_0]$ , one demands that the entropy  $S[p(z)]$  is equal to 1 [31]. This, together with the symmetry in  $-1 \leftrightarrow +1$  for  $x$  and  $z$ , constrains the form of the initial distribution to be parameterized as

$$p_0(x, z) = \left( \frac{1 + \mu_0}{4}, \frac{1 - \mu_0}{4}, \frac{1 - \mu_0}{4}, \frac{1 + \mu_0}{4} \right). \quad (\text{F1})$$

The probability  $p(x', z)$  then reads

$$p(x', z) = \left( \frac{r + e^{-2tu} r - 2u + e^{-rt} (-r + r\mu_0 - 2u\mu_0)}{4(r-2u)}, \frac{r - e^{-2tu} r - 2u + e^{-rt} (+r - r\mu_0 + 2u\mu_0)}{4(r-2u)}, \frac{r - e^{-2tu} r - 2u + e^{-rt} (+r - r\mu_0 + 2u\mu_0)}{4(r-2u)}, \frac{r + e^{-2tu} r - 2u + e^{-rt} (-r + r\mu_0 - 2u\mu_0)}{4(r-2u)} \right), \quad (\text{F2})$$

and we can explicitly compute the mutual information:

$$\begin{aligned} I[x_t, z_0] &= \frac{1}{2} \left( 1 + \mu_0 e^{-rt} + \frac{r}{r-2u} (e^{-2ut} - e^{-rt}) \right) \log_2 \left[ 1 + \mu_0 e^{-rt} + \frac{r}{r-2u} (e^{-2ut} - e^{-rt}) \right] + \\ &+ \frac{1}{2} \left( 1 - \mu_0 e^{-rt} + \frac{r}{r-2u} (e^{-2ut} - e^{-rt}) \right) \log_2 \left[ 1 - \mu_0 e^{-rt} + \frac{r}{r-2u} (e^{-2ut} - e^{-rt}) \right]. \end{aligned} \quad (\text{F3})$$

The above equation can again be rewritten as in Eqn. 3 of the paper if we introduce

$$\mu = \mu_0 e^{-rt} + \frac{r}{r-2u} (e^{-2ut} - e^{-rt}). \quad (\text{F4})$$

For  $t=0$ ,  $\mu = \mu_0$  and the information is maximized by  $\mu = \mu_0 = \pm 1$ .

## Appendix G: Models $B$ and $\tilde{B}$

Models  $B$  and  $\tilde{B}$  are extensions of model  $A$  and  $\tilde{A}$ , respectively, where we allow  $z$  to also be a repressor of  $x$

( $s_m \neq 0, s_p \neq 0$ , and we no longer demand  $r_p = r_m$ ). As in models  $A$  and  $\tilde{A}$ , we do not allow feedback from  $x$  to  $z$ , meaning that the transition rates for  $z$  do not depend on the state of  $x$  (i.e.  $u_m = u_p = u$  and  $d_p = d_m = d$ ). The results for the optimal parameters and topologies of these models are very similar to models  $A$  and  $\tilde{A}$  and are not shown in the main text. We plot the corresponding results in Fig. 6 here. We note that in model  $\tilde{B}$  we have absorbing states, similar to those of model  $\tilde{C}$  shown in the main text. As in model  $\tilde{C}$ , the symmetries of the problem allow many equivalent (under permutation of labels) parameter settings: we show all of them in Fig. 7.

However, unlike in model  $\tilde{C}$  where all states are visited, in each optimal setting of model  $\tilde{B}$  one state (shadowed in Fig. 7) is never visited.

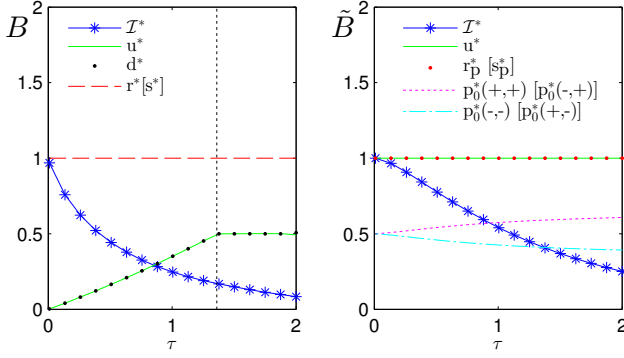


FIG. 6: Example of optimal parameters, along with transmitted information  $\mathcal{I}^*$ , for models  $B$  and  $\tilde{B}$ . Parameters in square brackets refer to alternative optimal topologies where  $z$  down-regulates  $x$  instead of up-regulating it (see Fig. 7  $B$  and  $\tilde{B}$ ). Subscripts  $m, p$  are omitted when  $x_m = x_p$  (with  $x = u, d, r, s$ ). In model  $\tilde{B}$  the optimal rates do not change with  $\tau$  and the optimal initial distribution is split slightly unevenly between the beginning and end state.

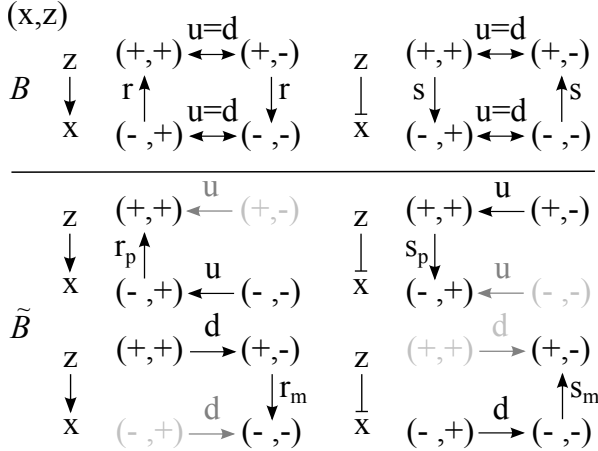


FIG. 7: Full set of optimal topologies and dynamics for models  $B$  and  $\tilde{B}$ . The topologies of these optimal networks were found by inspecting the optimal rates manually. In model  $\tilde{B}$  some states (shown in gray in the figure) are never visited.

#### Appendix H: Model $\tilde{C}$

Model  $\tilde{C}$  is the same as model  $C$ , except for the fact that the steady state assumption is now relaxed and the system is optimized also over the initial distribution (as in models  $\tilde{A}$  and  $\tilde{B}$ ). The results are discussed in detail in the main text; here, we simply show in Fig. 8 the four maximally informative topologies that have been found: each of them can be labeled as a “push-pull” network and features an absorbing state.

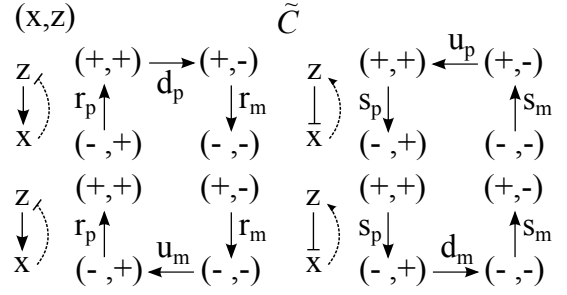


FIG. 8: Optimal topologies and dynamics for model  $\tilde{C}$ , where the steady state assumption is relaxed and feedback is present. The topologies of these optimal networks were found by inspecting the optimal rates manually. The optimal initial distribution is split slightly unevenly between the beginning and end state.

#### Appendix I: Information decay in a system with fixed parameters.

In the main text, for each relaxation time, we optimize the parameters of the system in order to get a maximally informative network. However, real regulatory networks have a fixed set of parameters for all times. In Fig. 9 we thus plot the mutual information between the initial input  $z_0$  and a delayed output  $x_t$  as a function of the delay  $t$ , for different sets of fixed parameters. In particular we consider models  $A$ ,  $C$  and  $\tilde{A}$ . We see that information decays more slowly if the system is initially in steady state, and therefore continues being in steady state (models  $A$  and  $C$ ). However, being out of steady states (model  $\tilde{A}$ ) allows for larger initial information transmission for small times.

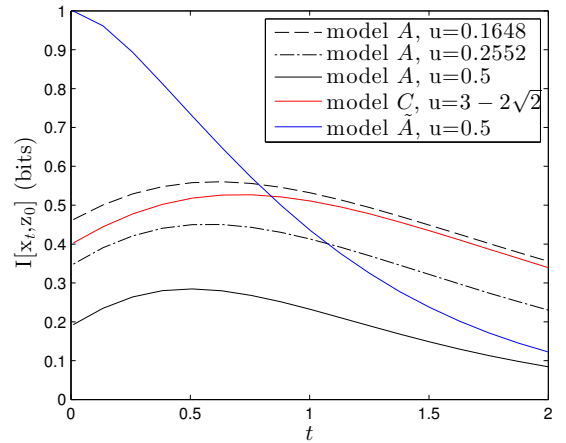


FIG. 9: The mutual information between the initial input  $z_0$  and a delayed output  $x_t$  as a function of the delay  $t$  for different sets of fixed parameters. A comparison of model  $A$  (black lines), model  $C$  (red line) and model  $\tilde{A}$  (blue line).

- 
- [1] G. Tkačik and A. M. Walczak, J. of Phys.: Cond. Mat. **23** (2011).
  - [2] A. M. Walczak, A. Mugler, and C. H. Wiggins, PNAS **105** (2009).
  - [3] R. Cheong et al., Science **334** (2011).
  - [4] E. Emberly, Phys. Rev. E **77** (2008).
  - [5] E. Ziv, I. Nemenman, and C. H. Wiggins, PLoS One **2** (2007).
  - [6] W. de Ronde, F. Tostevin, and P. R. ten Wolde, Phys. Rev. E **86** (2012).
  - [7] A. Mugler, A. M. Walczak, and C. H. Wiggins, Phys. Rev. Lett. **105** (2010).
  - [8] F. Tostevin and P. R. ten Wolde, Phys. Rev. E **81** (2010).
  - [9] F. Tostevin and P. R. ten Wolde, Phys. Rev. Lett. **102** (2009).
  - [10] W. de Ronde, F. Tostevin, and P. R. ten Wolde, Phys. Rev. E **82** (2010).
  - [11] B. Alberts et al., *Essential Cell Biology* (Garland Science, 2010), iii ed.
  - [12] J. E. Segall, S. M. Block, and H. C. Berg, PNAS **83** (1986).
  - [13] D. Fuller et al., PNAS **107** (2010).
  - [14] T. T. Le et al., Biophys. J. **90** (2006).
  - [15] E. Guisbert et al., Genes Dev. **18** (2004).
  - [16] G. Lahav et al., Nat. Gen. **36** (2004).
  - [17] A. R. Brasier, Cardiovasc Toxicol. **6** (2006).
  - [18] I. Nemenman, Phys. Biol. **9** (2012).
  - [19] U. Alon, NRG **8** (2007).
  - [20] G. Tkačik, A. M. Walczak, and W. Bialek, Phys. Rev. E **80** (2009).
  - [21] G. Tkačik, A. M. Walczak, and W. Bialek, Phys. Rev. E **85** (2012).
  - [22] J. Stricker et al., Nature. **456** (2002).
  - [23] Code that performs the optimization is available at <http://infodyn.sourceforge.net>.
  - [24] For  $r=2u$ ,  $\mu=e^{-rt}/2(1-2rt)$ .
  - [25] Note that, in un-rescaled time, maximal information transmission occurs for  $u \rightarrow 0$  with  $r$  finite, for any  $t > 0$ , as  $I[x_t, z_0] \rightarrow 1$ .
  - [26] We remove arbitrariness in the choice of time units by fixing the magnitude of the maximum rate to be 1.
  - [27] Delay in mRNA production was shown to be a necessary element of stable oscillations, making them hard to observe in synthetic networks [22].
  - [28] The case of a steady state initial distribution is a special case of a system that enjoys this property.
  - [29] In each of the 4 degenerate topologies, a different states becomes absorbing.
  - [30] Such design is reminiscent of Shannon's *Ultimate Machine*: a device with a switch that, when it is turned on, activates an arm that switches it off.
  - [31] For any distribution  $p(a, b)$ ,  $I[a, b] < \min\{S[a], S[b]\}$ . In this case, if we want  $I[x_t, z_0]$  to reach 1 bit, it must be the case that  $S[x_t]$  and  $S[z_0]$  are both 1 bit.